

How To Use Propensity Score Analysis

Lisa Kaltenbach, MS
Department of Biostatistics
lisa.kaltenbach@vanderbilt.edu
April 11, 2008

Outline

- Background/Motivation
- Propensity Score Estimation
- Propensity Score Matching
- Regression Adjustment/Stratification
- Example Code/Analysis
- Conclusions
- References

Motivational Anecdote

- Two heart surgeons walk into a room.
 - The first surgeon says, “Man, I just finished my 100th heart surgery!”.
 - The second surgeon replies, “Oh yeah, I finished my 100th heart surgery last week. I bet I'm a better surgeon than you. How many of your patients died within 3 months of surgery? Only 10 of my patients died.”
 - First surgeon smugly responds, “Only 5 of mine died, so I must be the better surgeon.”
 - Second surgeon says, “My patients were probably older and had a higher risk than your patients.”

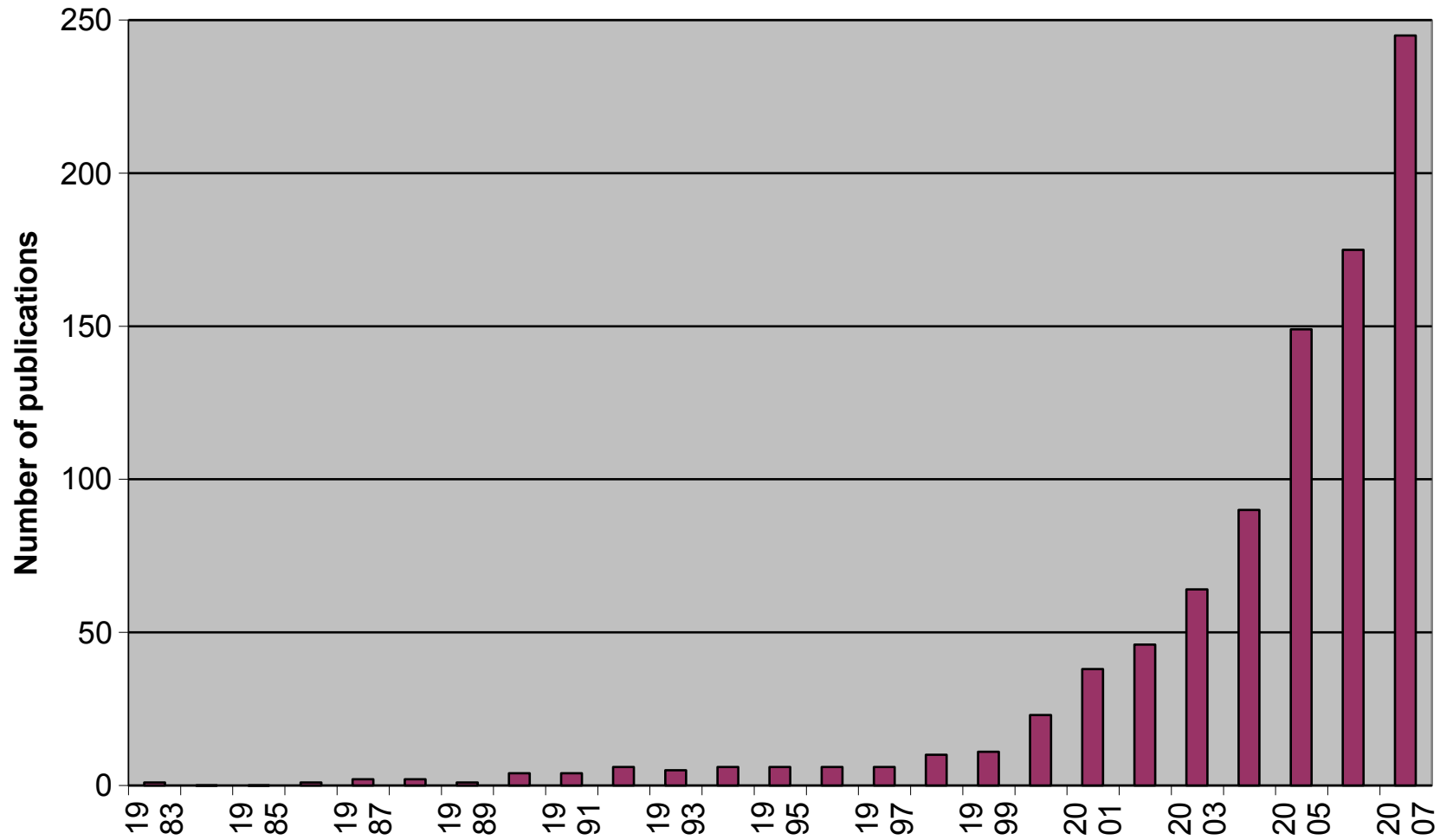
Comparing apples to oranges?

- There may be important differences in patient characteristics between treatment groups.
- Want to show difference in outcome is attributable to difference in treatment (or patient condition) and not due to comparing apples to oranges.
- Nonrandomized comparisons give rise to apples-and-oranges scepticism.
- Sometimes it is infeasible or unethical to assign patients to different treatments.

Purpose of Propensity Scores

- Can produce apples-to-apples comparison under some nonrandomized conditions.
- Provides a way to summarize covariate information about treatment selection into a scalar value.
- Can be used to adjust for differences via study design (*matching*) or during estimation of treatment effect (*stratification/regression*).
- Analysis limitations: <10 events/variable (EPV), estimated regression coefficients may be biased & SE's may be incorrect (Peduzzi *et al*, 1996; Harrell *et al*, 1985).

Publications in Pub Med with phrase "Propensity Score"



Notation/Definition

- Treatment Groups (E):
 - Let E+ denote group with exposure.
 - Let E- denote group without exposure.
- Disease Outcome (D):
 - Let D+ denote group with disease outcome.
 - Let D- denote group without disease outcome.
- Propensity score (PS):
 - For an individual is the conditional probability of being treated given the individual covariates.
 - $PS = \text{Estimated Pr}(E+ | \text{covariates})$.

Propensity Score Estimation

- Identify potential confounders.
 - Current convention: If uncertain whether a covariate is a confounder, include it.
- Model E+ (typically dichotomous) as a function of covariates using entire cohort:
 - E+ is outcome for propensity score estimation.
 - Do not include D+.
 - Logistic regression typically used.
 - Propensity Score = estimated $\Pr(E+ | \text{covariates})$.
 - Can use PS as a continuous variable or create quantiles.

Natural Question

- Why estimate $\Pr(E+ | \text{covariates})$ when we already know $E+$?
 - Adjusting observed $E+$ with probability of $E+$ (“propensity”) creates a “quasi-randomized” experiment.
 - For $E+$ & $E-$ patients with same propensity score, can imagine that they were “randomly” assigned to each group.
 - Subjects in $E+/E-$ groups with nearly equal propensity scores tend to have similar distributions in covariates used to estimate propensity.

A Balancing Score

- For a given propensity score, one gets unbiased estimates of average E+ effect.
- Can include a large number of covariates for PS estimation. Original paper applied PS methodology to observational study comparing CABG to medical treatment, adjusting for 74 covariates in the PS model.
- Want to assess adequacy of propensity score to adjust for effects of covariates by testing for differences in individual covariates between E+ & E- after adjusting for propensity score (often we stratify by propensity score quantiles) .

Applications

- Matching.
- Regression adjustment/stratification.
- Weighting (each patient's contribution to regression model).
 - Inverse-probability-of-tx-weighted see Robin *et al*, 2000.
 - Standardized mortality ratio-weighted estimator see Sato *et al*, 2003.

Propensity Score Matching

- Match on a single summary measure.
 - Consider study on low-dose aspirin & mortality
 - Age is a strong confounder, but can be controlled by matching.
 - Extending this to many factors becomes cumbersome quickly.
- Useful for studies with limited number of E+ patients and a larger number of E- patients and need to collect additional measures (ex., blood samples).

Matching Techniques

- Nearest available matching on estimated propensity score:
 - Select E+ subject.
 - Find E- subject with closest propensity score,
 - Repeat until all E+ subjects are matched.
 - Easiest method in terms of computational considerations.
- Others:
 - Mahalanobis metric matching (uses propensity score & individual covariate values.
 - Nearest available Mahalanobis metric matching w/ propensity score-based calipers.

Matching in R

- Install the “Matching” package by Jasjeet Sekhon.
- <http://sekhon.berkeley.edu/matching/>
- Match(): performs multivariate and PS matching.
- MatchBalance(): provides a variety of univariate tests to determine if balance exists.
- Matchby(): is a wrapper for Match() function which separates the matching problem into subgroups defined by a factor.

Matching in Stata

- Install psmatch2 package created by Edwin Leuven and Barbara Sianesi.
- <http://ideas.repec.org/c/boc/bocode/s432001.html>
- psmatch: implements various types of propensity score matching estimators.
 - one-to-one, k -nearest neighbors, radius, kernel, local linear regression, spline, Mahalanobis.

Matching in SAS

- Macros created and maintained by statisticians at the Mayo Clinic.
- Can be downloaded for free at:
 - <http://ndc.mayo.edu/mayo/research/biostat/sasmacros.cfm>
- *gmatch*: Computerized matching of cases to controls using the greedy matching algorithm with a fixed number of controls per case.
- *vmatch*: Computerized matching of cases to controls using variable optimal matching.
 - The number of controls per case is allowed to vary with only the total fixed.
 - Controls may be matched to cases using one or more factors (X's).

Example: The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients (JAMA 1996; 276: 889-897)

- *Objective:* Examine association between RHC use during 1st 24 hours of ICU care & subsequent survival, length of stay, intensity of care, & cost of care.
- *Design:* Prospective Cohort study.
- *Setting:* 5 US teaching hospitals from 1989 through 1994.
- *Subjects:* Critically ill adult patients receiving care for 1 of 9 prespecified disease categories (acute respiratory failure, COPD, CHF, cirrhosis, nontraumatic coma, colon cancer metastatic to the liver, non-small cell cancer of the lung, multiorgan system failure with malignancy or sepsis).
- *Exposure:* RHC+/RHC- (at discretion of physician & thus may be confounded with patient factors related to the outcome).
- *Disease:* Survival, cost of care, intensity of care, length of stay in ICU & hospital.
- *Analysis:* McNemar's, linear regression, Cox proportional hazards.

PS Estimation: $\Pr(\text{RHC+} \mid \text{covariates})$

- Choice of RHC+/RHC- was at the discretion of physician.
- Treatment selection may be confounded with patient characteristics related to the outcome.
 - For example, patients with low BP may be more likely to receive RHC and may be more likely to die.
- Panel of 7 specialists in critical care specified variables related to RHC use.
- Use logistic regression with outcome RHC.

PS Estimation: $\Pr(\text{RHC+} \mid \text{covariates})$ [2]

- Covariates included:
 - age, sex, race, yrs of education, income, medical insurance, primary and secondary disease category, admission diagnosis, ADL and DASI, DNR status, cancer, 2-month survival probability, acute physiology component of APACHE III score, Glago Coma Score, weight, temperature, BP, respiratory rate, heart rate, PaO₂/FiO₂, PaCO₂, pH, WBC count, hematocrit, sodium, potassium, creatinine, bilirubin, albumin, urine output, comorbid illness.
- Adequacy of PS to adjust for effects of covariates assessed by testing for differences in individual covariates between RHC+/RHC- patients after stratifying by PS quintiles.
 - Model each covariate as a function of RHC & PS quintiles.
 - No detectable imbalances if not related RHC after PS adjustment.

Propensity Score Matching

- Each RHC+ was matched with a RHC- with same disease category & closest PS (± 0.03).
- Difference in PS for each pair calculated, & each pair with a positive difference matched with pair having negative difference closest in magnitude.
 - Assures equal numbers of pairs with positive & negative propensity differences.
- 1008 matched pairs.

Propensity-matched analysis of RHC & survival

Survival Interval	Survival, n (%)		OR (95% CI)
	RHC-	RHC+	
30day	677(67.2)	630 (62.5)	1.24 (1.03 – 1.49)
60 day	604 (59.9)	550 (54.6)	1.26 (1.05 – 1.52)
180 day	522 (51.2)	464 (46.0)	1.27 (1.06 – 1.52)
Hospital	629 (63.4)	565 (56.1)	1.39 (1.15 – 1.67)

Regression

Adjustment/Stratification

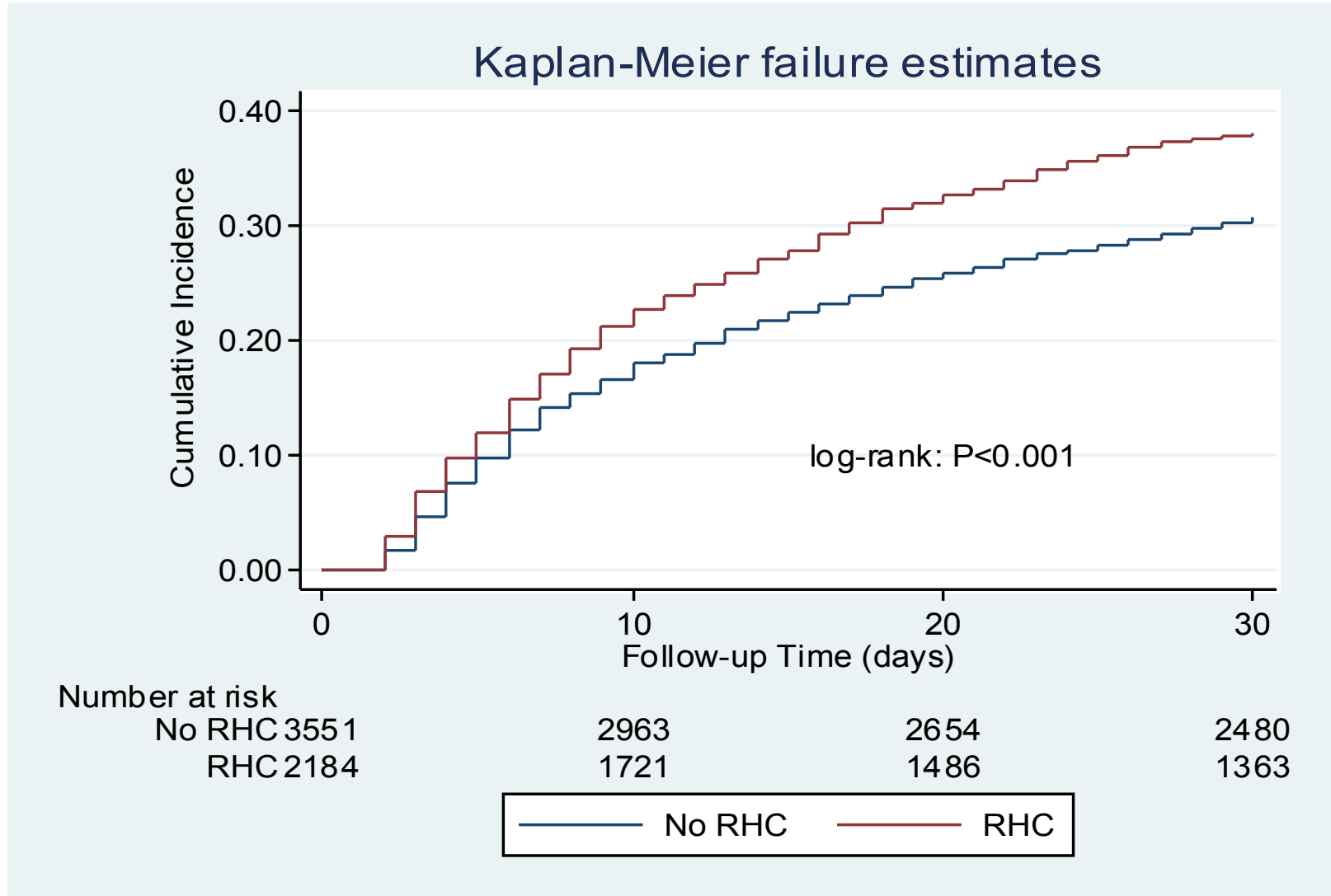
- Can include PS in final analysis model as a continuous measure or create quantiles and stratify.
- Rosenbaum & Rubin (1983) showed that perfect stratification based on PS will produce strata where average tx effect within strata is an unbiased estimate of the true tx effect.

Example: The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients (JAMA 1996; 276: 889-897)

- Dataset and description of dataset are available online at <http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/DataSets> : rhc.*.
- *Objective:* Examine association between RHC use during 1st 24 hours of ICU care & subsequent survival, length of stay, intensity of care, & cost of care.
- *Design:* Prospective Cohort study.
- *Setting:* 5 US teaching hospitals from 1989 through 1994.
- *Subjects:* Critically ill adult patients receiving care for 1 of 9 prespecified disease categories (acute respiratory failure, COPD, CHF, cirrhosis, nontraumatic coma, colon cancer metastatic to the liver, non-small cell cancer of the lung, multiorgan system failure with malignancy or sepsis).
- *Exposure:* RHC+/RHC- (at discretion of physician & thus may be confounded with patient factors related to the outcome).
 - swang1 (1=RHC+, 0=RHC-).
- *Disease:* Survival
 - t3d30 = time-to-death; censor (1=died, 0=censored).
- *Analysis:* Cox proportional hazards.

Kaplan-Meier plot by RHC status Stata Code:

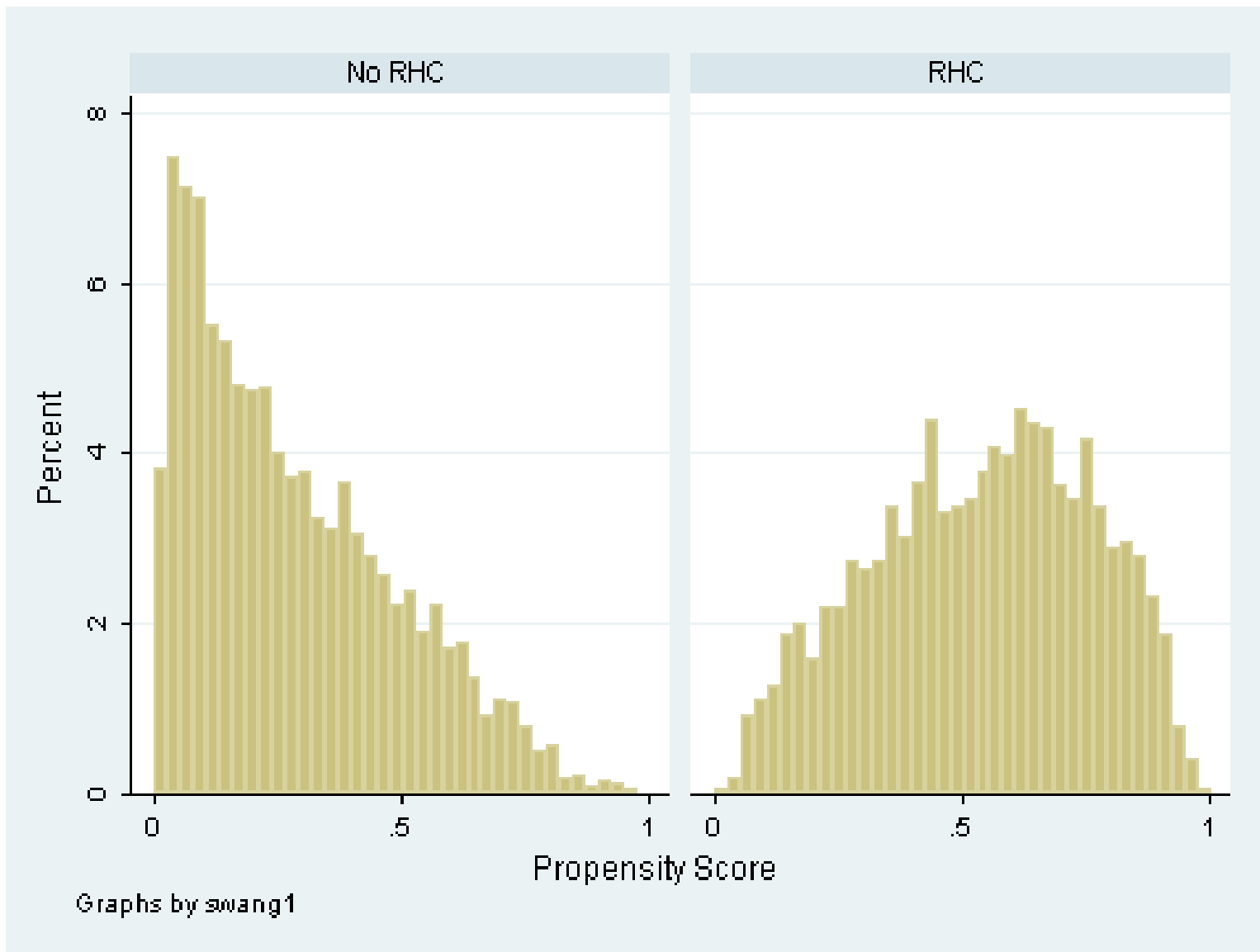
```
sts graph, failure by(swang1) risktable ytitle(Cumulative Incidence)  
ylab(0(0.1)0.4,angle(horizontal)) xtitle(Follow-up Time (days))  
text(0.1 20 "log-rank: P<0.001")
```



Propensity Score Model

- Logistic regression: RHC+/RHC- dependent variable & adjust for 50 risk factors (selected by a panel of 7 specialists in critical care).
 - xi: logistic i.swang1 age i.sex i.race edu i.income i.ninsclas i.cat1 das2d3pc i.dnr1 i.ca surv2md1 aps1 scoma1 wtkilo1 temp1 meanbp1 resp1 hrt1 pafi1 paco21 ph1 wb1c1 hema1 sod1 pot1 crea1 bili1 alb1 i.resp i.card i.neuro i.gastr i.renal i.meta i.hema i.seps i.trauma i.ortho i.cardiohx i.chfhx i.dementhx i.psychhx i.chrpulhx i.renalhx i.liverhx i.gibledhx i.malighx i.immunhx i.transhx i.amihx
 - predict prop_score

Propensity Score Distribution



Covariates related to RHC after ps adjustment (selected risk factors)?

****Create PS quintiles

```
xtile ps_quintiles = prop_score, nq(5)
```

****Assess PS – adjusted age

```
xi: regress age i.swang1 i.ps_quintiles
```

****Assess PS – adjusted gender

```
xi: logistic gender i.swang1 i.ps_quintiles
```

Covariates related to RHC after PS adjustment? [2]

	PS - adjusted, p-value	
	Before	After
Age	0.026	0.945
Gender	0.001	0.731
APACHE score	<0.001	0.100
Weight (kg)	<0.001	0.53
Mean BP	<0.001	0.255
Respiratory Rate	<0.001	0.531
WBC	0.002	0.604
Creatinine	<0.001	0.470

RHC & survival

*** unadjusted model

xi: stcox i.swang1

*** fully-adjusted model

xi: stcox i.swang1 age i.sex i.race edu i.income i.ninsclas i.cat1
das2d3pc i.dnr1 i.ca surv2md1 aps1 scoma1 wtkilo1 temp1
meanbp1 resp1 hrt1 pafi1 paco21 ph1 wblc1 hema1 sod1
pot1 crea1 bili1 alb1 i.resp i.card i.neuro i.gastr i.renal i.meta
i.hema i.seps i.trauma i.ortho i.cardiohx i.chfhx i.dementhx
i.psychhx i.chrpulhx i.renalhx i.liverhx i.gibledhx i.malighx
i.immunhx i.transhx i.amihx

*** propensity score (linear) model

xi: stcox i.swang1 prop_score

*** propensity score (quintiles) model

xi: stcox i.swang1 i.ps_quintiles

Results

Model	HR (95% CI)
Unadjusted	1.30 (1.19 – 1.43)
Multivariable	1.24 (1.12 – 1.38)
PS-adjusted (linear term)	1.22 (1.10 – 1.36)
PS-adjusted (quintiles)	1.24 (1.11 – 1.37)

Note: 1918 deaths & 50 covariates (excluding RHC) yields 40 events/confounder.

Conclusions

- Benefits:
 - Useful when adjusting for a large number of risk factors & small number of events per variable.
 - Useful for matched designs (saving time & money).
- Limitations:
 - Can only adjust for observed covariates.
 - PS methods work better in larger samples to attain distributional balance of observed covariates.
 - Bias may occur.
 - Including irrelevant covariates in propensity model may reduce efficiency.

Thanks!

- Patrick Arbogast, PhD.
- All of you for coming.

References

- Blackstone EH. Comparing apples and oranges. *J Thoracic and Cardiovascular Surgery* 2002; **1**: 8-15.
- Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003; **158**: 280-287.
- Connors Jr AF, Speroff T, Dawson NV, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. *JAMA* 1996; **276**: 889-897.
- D'Agostino Jr RB, Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17**: 2265-2281.
- Harrell FE, Lee KL, Matchar DB, Reichart TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treatment Reports* 1985; **69**: 1071- 1077.
- Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006; **163**: 262-270.
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**: 1373-1379.
- Robbins JM, Hernan MA, Brumback B. Marginal structural models and causal inferences in epidemiology. *Epidemiology* 2000; **11**: 550-560.
- Rosenbaum PR. *Observational Studies*. New York, NY: Springer-Verlag, 2002.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41-55.
- Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annal of Internal Medicine* 1997; **127**: 757-763.
- Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003; **14**: 680-686.